Technical Report ARL-71-21/AFOSR-71-7
October 1971

Research supported by:

Air Force Office of Scientific Research
Contract F44620-70-C-0105

RESPONSE SURFACE METHODOLOGY DESIGN VARIANTS USEFUL IN HUMAN PERFORMANCE RESEARCH

Christine Clark

Robert C. Williges

Presented at 15th Annual Meeting
of the Human Factors Society
New York, New York

University of Illinois at Urbana-Champaign
AVIATION RESEARCH LABORATORY
INSTITUTE OF AVIATION
University of Illinois-Willard Airport
Savoy, Illinois

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Aviation Research Laboratory, Institute of Aviation, University of Illinois, Urbana, Illinois	2a. REPORT SECURITY CLASSIFICATION Unclassified
	2b. GROUP

3. REPORT TITLE

RESPONSE SURFACE METHODOLOGY DESIGN VARIANTS USEFUL IN HUMAN PERFORMANCE RESEARCH

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Scientific Interim

5. AUTHOR(S) (First name, middle initial, last name)

Christine Clark
Robert C. Williges

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
October 1971	23	9

8a. CONTRACT OR GRANT NO. F44620-70-C-0105	9a. ORIGINATOR'S REPORT NUMBER(S) ARL-71-21/AFOSR-71-7
b. PROJECT NO. 9778	
c. 61102F	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFOSR - TR - 72 - 1667
d. 681313	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
TECH,OTHER	Air Force Office of Scientific Research (NL), 1400 Wilson Boulevard, Arlington, Virginia 22209

13. ABSTRACT

Selected Response Surface Methodology (RSM) designs that are viable alternatives in human performance research are discussed. Two major RSM designs that are variations of the basic, blocked, central-composite design have been selected for consideration: 1) central-composite designs with multiple observations at only the center point, 2) central-composite designs with multiple observations at each experimental point. Designs of the latter type are further categorized as: a) designs which collapse data across all observations at the same experimental point; b) between-subjects designs in which no subject is observed more than once, and observations at each experimental point may be multiple and unequal or multiple and equal; and c) within-subject designs in which each subject is observed only once at each experimental point. The ramifications of these designs are discussed in terms of various criteria such as rotatability, orthogonal blocking, and estimates of error.

1a

DD FORM 1 NOV 65 1473

Unclassified
Security Classification

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Response Surface Methodology						
Computer Program						
Statistical Analyses						
Experimental Design						
Multiple Regression						
Correlation						
Analysis of Variance						

Unclassified
Security Classification

RESPONSE SURFACE METHODOLOGY DESIGN VARIANTS USEFUL IN HUMAN PERFORMANCE RESEARCH

Christine Clark
Robert C. Williges

i c

FOREWORD

This paper represents the theoretical basis for a research project currently being conducted by the Aviation Research Laboratory of the Institute of Aviation, University of Illinois at Urbana-Champaign. This research is being supported by the Life Sciences Program, Air Force Office of Scientific Research. This project is one of eight tasks, four of which are in the area of general pilot selection, training, and performance and four of which deal with avionics system design principles.

INTRODUCTION

Although Response Surface Methodology (RSM) was originally developed by Box and Wilson (1951) for use in chemical research, its utility in psychological research, especially in studies of human performance, also has been documented (Meyer, 1963; Simon, 1970; Williges & Simon, 1971). RSM is a tool for assessing the effect of continuous, quantitative factors upon some response of interest to the investigator. Unlike traditional factorial, analysis of variance designs, RSM focuses primarily on the functional relationship which exists between the response and the factors, rather than merely determining the significance of various factors. Indeed, if one's aim is to determine the quantitative relationship between human performance and a host of equipment parameters, the traditional methods fall short. In a recent application of RSM to a target recognition experiment, Simon (1970) noted several direct benefits which accrue from a RSM regression equation approximating the response surface. Such equations allow one to estimate performance levels at given levels of system variables, to estimate necessary levels of system variables given a designated performance level, to determine the tradeoffs among system parameters while optimizing performance level, and to judge the relative importance of the various system parameters.

In addition to providing an approximation of the relationship between performance and factors, RSM advances a variety of experimental designs to achieve that estimate as efficiently and as economically as possible. When using factorial designs, the investigator is often forced by practical considerations to limit the number of factors studied to even less than the number that he believes has a critical effect on performance. In such a case he must conduct multiple studies yielding isolated investigations of only a few factors at any one time. This results in an unrealistic view of a system in which a number of factors operate simultaneously to influence performance. By allowing the investigator to consider large numbers of factors within a single study, RSM proves a valuable investigatory tool. Through strategic sampling of data points, RSM provides the most essential

information and allows one to decide whether or not the collection of additional data is merited.

Most RSM designs are special cases of the Box and Wilson (1951) central-composite design. Because these designs were originally developed for chemical research, it is not unreasonable to anticipate the need for some modification in their basic design to make the method more adaptable to research involving human subjects. The purpose of this paper is to suggest several appropriate design modifications that attempt to retain as many of the positive traits of the RSM technique as possible. Before discussing these modifications, a brief description of central-composite designs is necessary.

CENTRAL-COMPOSITE DESIGNS

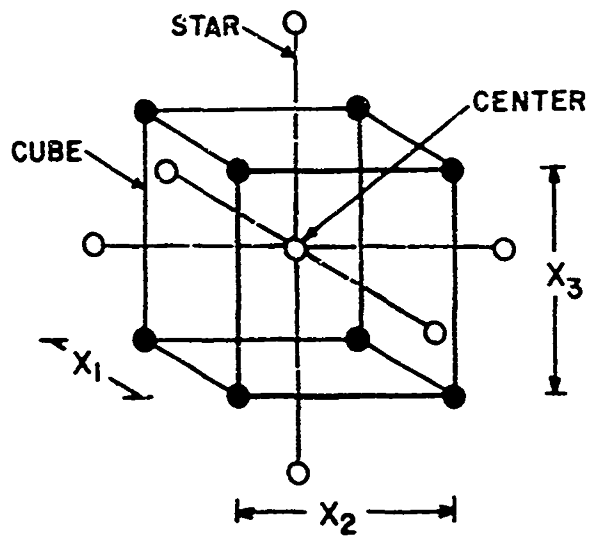
By way of example, suppose that an investigator is interested in predicting radar target detection, Y, as a function of display resolution, X_1, visual angle, X_2, and random noise, X_3. When developing an equation to approximate the relationship between a given response and a set of experimental variables, an investigator generally measures the desired response at relatively few data points, each defined by some unique combination of experimental variable levels. From the resulting data he then makes inferences about the nature of the underlying relationship. The basic central-composite design details a strategy to be followed in selecting these data collection points that insures orthogonality of factors while substantially limiting the required number of observations. The identifying feature of a central-composite design is that all the data collection points are spaced symmetrically about and at a specified distance from an arbitrary center point that is determined by the experimental region of greatest interest to the investigator. The data points can be thought of as sampling three portions of the overall experimental region. In a K-factor design, a 2^K factorial (or fractional factorial) is combined with a 2K star plus an appropriate number of center points. For the hypothetical case of K=3 outlined above, the design can be diagrammed as in

Figure 1. Eight unique data points lie at the vertices of a cube; 6 other unique data points form a six arm star, radiating from yet a fifteenth unique point at the center of the cube and extending beyond the surfaces of the cube.

For each factor under consideration, five levels of it must occur. Furthermore, these five levels must correspond to the coded values 0, ± 1, and $\pm \alpha$, where α represents the length of the star arm or the radius of the hypersphere on which data are located. The appropriate real world values for each factor are determined by the range of values in which one is interested and by the constraints placed upon them by the required coded values. For example, suppose that display resolution ranges in value from 168 to 504 TV lines/dm and that these two limiting values are subsequently assigned coded values of $-\alpha = -1.682$ and $\alpha = +1.682$, respectively. Then the corresponding transformation $X^*_1 = \frac{X_1 - 336}{100}$ dictates that the other levels of X_1 be 236, 336, 436, which correspond to the coded values -1, 0, and 1, respectively. Although coding the appropriate raw score, independent variable levels is not necessary, the use of linear transformations of the data permits a simpler analysis without affecting the result of any subsequent statistical test. On occasion this rigid demand regarding the selection of data points makes the central-composite design impractical for some human factors studies. For example, variables such as target type, target complexity, and briefing instructions are not readily quantifiable. Moreover, it is sometimes neither practical nor feasible to measure even certain quantifiable variables at the five levels specified by the central-composite design.

Replication

The original central-composite design calls for repeated observations to be made at the designated center point only. As noted by Cochran and Cox (1957), these replicated center points provide additional degrees of freedom to permit an estimate of experimental error, and they determine the precision of the estimate at and near the center. Choosing the appropriate number of center points results in a design in which the standard error of estimate is roughly the same at all



OBSERVATION	X_1	X_2	X_3
1	1.0	-1.0	1.0
2	1.0	1.0	-1.0
3	-1.0	1.0	1.0
4	-1.0	-1.0	-1.0
5	-1.0	1.0	-1.0
6	-1.0	-1.0	1.0
7	1.0	-1.0	-1.0
8	1.0	1.0	1.0
9	-1.682	0.0	0.0
10	0.0	-1.682	0.0
11	0.0	0.0	-1.682
12	1.682	0.0	0.0
13	0.0	1.682	0.0
14	0.0	0.0	1.682
15	0.0	0.0	0.0
16	0.0	0.0	0.0
17	0.0	0.0	0.0
18	0.0	0.0	0.0
19	0.0	0.0	0.0
20	0.0	0.0	0.0

Figure 1. Second-order, rotatable, central-composite design in three variables with coded value coordinates of data points.

points within the experimental region. Hence, the estimate of error computed at the center can be used as an estimate of error throughout the experimental K-space, thereby minimizing redundancy. Too many replications at the center yield standard errors of estimate which increase rapidly for those points farther from the center. On the other hand, with too few replications of the center point, the standard error is apt to be greater at the center than at the surrounding data points.

Rotatability

Another desirable property of many central-composite designs is rotatability (Box and Hunter, 1957). Rotatability insures equal reliability of predicted responses at all data points equidistant from the center. This is an especially convenient design quality in exploratory work when the investigator is ignorant of the response surface and its relative orientation to the orthogonal factor axes. Rotatability imposes the additional constraint that the value of α be equal to $2^{K/4}$. As the number of factors increases , fractional replicates of the full 2^K factorial are used, thereby reducing the overall size of the experiment. In general, when fractional factorial designs of $(1/2)^p$ are used in place of the full factorial, then α must equal $2^{(k-p)/4}$ if rotatability is to exist (Cochran and Cox, 1957). Thus, if the hypothetical three factor design diagrammed in Figure 1 is to be rotatable, the star arm must be 1.682 (coded value). To insure roughly equal precision of prediction across the entire experimental region, the center point is replicated 6 times. When complete, the design involves a total of 20 observations with 14 of the experimental points lying on the surface of a sphere of radius 1.682, and with six observations being made at the center point (0, 0, 0).

Blocking

An additional feature of RSM designs that affords the investigator greater efficiency and flexibility is blocking. Under blocking conditions subsets of the total data collection points are studied together. If the blocking is orthogonal, any differences in mean performance among blocks are independent of any main effects due to the independent variable manipulations, and as such, they do not

affect the underlying quantitative relationship between factors and performance. This aspect of design is valuable to the human factors engineer who is concerned with isolating potential effects due to such factors as different experimenters, changes in apparatus, and environmental conditions. Recall the investigator studying radar target detection latency as affected by display resolution, visual angle, and random noise. It is unlikely that all the necessary data can be collected during a single flight or perhaps not even in the same aircraft. By taking advantage of orthogonal blocking techniques, he can guard against the parameters of the RSM prediction equation being affected by such differences.

Orthogonal blocking places additional constraints on the central-composite design concerning the selection of α and the number of center points. These parameters must be chosen to insure that the average predicted response level is the same for every block. Orthogonal blocking is guaranteed when the following condition is met (Box and Hunter, 1957):

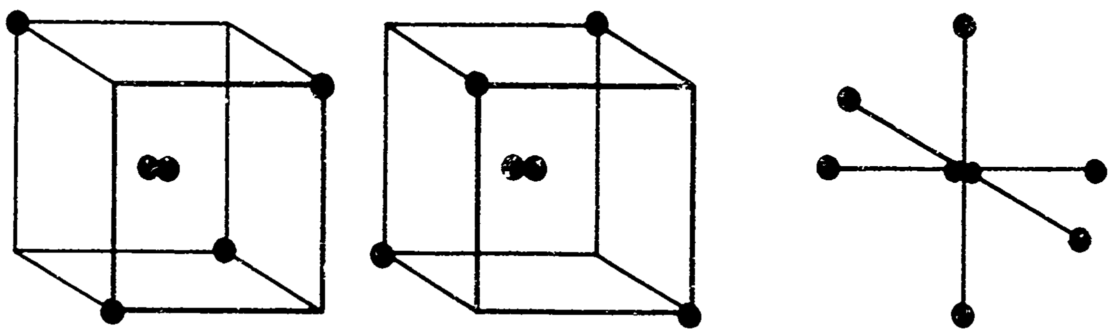
$$2^K/2\alpha^2 = (N_c + N_{co})/(N_s + N_{so}) \qquad (1)$$

where N_{co} and N_{so} are the number of center points added to the intact 2^K factorial portion and the 2 K star portion of the design, respectively. N_c and N_s reflect the number of noncenter points in the 2^K factorial and in the 2K star, respectively. Given the proposed design in Figure 1 for studying detection latency, orthogonal blocking can be achieved by dividing the 20 data points into subsets of 6, 6, and 8 observations, as depicted in Figure 2. The corresponding value of α is 1.633.

$$2^3/2\alpha^2 = (8+4)/(6+2)$$

$$\alpha = 1.633$$

Recalculating the real world levels of display resolution using this new α value, the investigator learns that the levels appropriate to the orthogonally blocked design are 173, 236, 336, 436, 499. He must also recompute the appropriate real world levels of visual angle and random noise in a like manner. Note that



BLOCK 1	BLOCK 2	BLOCK 3
(1, -1, 1)	(-1, 1, -1)	(-1.633, 0, 0)
(1, 1, -1)	(-1, -1, 1)	(0, -1.633, 0)
(-1, 1, 1)	(1, -1, -1)	(0, 0, -1.633)
(-1, -1, -1)	(1, 1, 1)	(1.633, 0, 0)
(0, 0, 0)	(0, 0, 0)	(0, 1.633, 0)
(0, 0, 0)	(0, 0, 0)	(0, 0, 1.633)
		(0, 0, 0)
		(0, 0, 0)

Figure 2. Orthogonal blocking of second-order central-composite design in three variables with coded value coordinates of data points.

the value of α required to insure orthogonality is slightly different from the 1.682 value required for rotatability. In order to achieve orthogonal blocking, it is often necessary to sacrifice rotatability. In human factors applications, however, the potential gains from orthogonal blocking probably outweigh the risk of forfeiting rotatability.

Added flexibility can accrue from use of blocking techniques, as Box and Hunter (1957) illustrated when they employed blocking to facilitate exploration of a response surface. A properly blocked design permits research to be conducted in stages. Each block of data points from the complete second-order design constitutes a first-order, rotatable central-composite design. Gathering data from the first series of blocks, the investigator can judge, for example, whether any of the original experimental variables merits being dropped from further consideration, or whether greater than a linear polynomial is needed to explain adequately the data. The ability to make such decisions at an early stage may mean that the investigator is able to conclude his study after collection of considerably less data than he had originally anticipated.

Analyses

Given data from a central-composite RSM design, the analyses of interest are:

1) Multiple regression analysis to determine a prediction equation which approximates the functional relationship between the desired response and the experimental factors. For the case of three independent variables, a general, complete second-order polynomial equation would be as follows:

$$Y = b_o + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_1^2 + b_5X_2^2 + b_6X_3^2 + b_7X_1X_2 + b_8X_1X_3 + b_9X_2X_3 \quad (2)$$

where Y is the resulting response level, X_1, X_2, and X_3 are levels of the three independent variables, and the b_i's are constant values known as partial regression coefficients. The multiple regression analysis determines the numerical values of those partial regression coefficients $b_o, \ldots, b_9$ according to a least squares criterion.

2) Analysis of variance (ANOVA) performed on the results of the preceding regression analysis. The ANOVA determines what proportion of the total variance in real world response is explained by the prediction equation; this estimate of the accuracy of the prediction equation is represented by the coefficient of determination, R^2. The ANOVA also determines the significance of each partial regression coefficient, indicating whether or not the corresponding variable is significant in determining the response level. In addition, the ANOVA can test for the significance of blocking, inter-subject differences, and lack of fit. A significant lack of fit term indicates that a higher order prediction equation, or a more complete equation of the same order, is required in order to represent the data adequately.

Consider again the study of radar target detection latency, Y, as a function of display resolution, visual angle, and random noise, X_1, X_2, and X_3, respectively. Hypothetical data for such a study are presented in Table 1. A multiple regression analysis of these hypothetical data yields the following generalized, first order prediction equation:

$$Y = 16.115 - 1.203\,X_1 - 0.503\,X_2 \div 0.847\,X_3.$$

Substituting given levels of the independent variables into this equation affords the investigator a corresponding predicted level of detection latency.

The results of a subsequent ANOVA performed on the regression analysis appear in Table 2. The derived equation accounts for nearly 74% of the total variance in detection latency. Each of the coefficients, excluding the constant term b_o, is significant at well beyond the .01 level. But, blocks are not significant. Noting that the lack of fit term is significant, the investigator will submit his data to a second multiple regression analysis to determine a higher order prediction equation.

TABLE 1

Hypothetical Data in Coded Form for a Three-Factor, Central-Composite RSM Design

Observation	Block	X_1 Resolution	X_2 Visual Angle	X_3 Random Noise	Y Detection Latency (Seconds)
1	1	1.0	-1.0	1.0	16.2
2	1	1.0	1.0	-1.0	14.3
3	1	-1.0	1.0	1.0	17.0
4	1	-1.0	-1.0	-1.0	17.4
5	1	0.0	0.0	0.0	15.5
6	1	0.0	0.0	0.0	15.8
7	2	-1.0	1.0	-1.0	16.8
8	2	-1.0	-1.0	1.0	18.1
9	2	1.0	-1.0	-1.0	14.9
10	2	1.0	1.0	1.0	16.2
11	2	0.0	0.0	0.0	15.0
12	2	0.0	0.0	0.0	14.8
13	3	-1.63	0.0	0.0	19.0
14	3	0.0	-1.63	0.0	17.3
15	3	0.0	0.0	-1.63	14.8
16	3	1.63	0.0	0.0	13.9
17	3	0.0	1.63	0.0	14.6
18	3	0.0	0.0	1.63	19.2
19	3	0.0	0.0	0.0	15.8
20	3	0.0	0.0	0.0	15.7

TABLE 2

First-Order Regression Analysis of Variance Summary Table
for Hypothetical Detection Latency Data

Source	Percent	df	Sum of Squares	Mean Square	F-Value
Regression	(73.92)	(3)	32.18	10.73	536.50 ***
b_1	44.25	1	19.26	19.26	963.00 ***
b_2	7.75	1	3.37	3.37	168.51 ***
b_3	21.93	1	9.54	9.54	477.00 ***
Residual	(26.08)	(16)	11.35	0.71	
Blocks	0.94	2	0.41	0.21	10.50 *
Lack of Fit	24.97	11	10.87	0.99	49.50 ***
Error	0.16	3	0.07	0.02	
Total	100.00	19	43.53		

Multiple Regression Coefficient = 0.86
Coefficient of Determination = 0.74

* $p < .05$
*** $p < .001$

DESIGN CONSIDERATIONS

In a recent article, Williges and Simon (1971) discussed several general advantages of the RSM technique which contribute to its potential value in human factors research. Among the most obvious benefits is the economy of data collection. Not only is sampling restricted to the experimental region of greatest interest, but also repeated observations are restricted to the center point of that region. As originally conceived, RSM was developed as a methodology for quickly locating optimums by means of a series of experiments each dependent on the results of the preceding one. More specifically, Box and Wilson (1951) were interested in determining the optimum combination of factor levels needed to produce the maximum yield from a chemical reaction. However, human factors engineers are largely interested in deriving global prediction equations which allow them to predict performance levels accurately throughout an entire range of factor levels.

When the goal is to approximate an entire response surface, rather than merely that portion of the surface surrounding the optimum, limiting multiple observations to a single experimental point may not be the most judicious strategy. Indeed, the actual variability in response may be so great across subjects and data points, that to presume the standard error of estimate at the center point is an adequate estimate of error at all points is unrealistic. A recent study concerning transfer of training (Baron and Williges, 1971) affords a striking demonstration of the effect of estimating experimental error at a single replicated point as opposed to estimating it across a series of replicated points. When replications were restricted to the center point, none of the experimental factors were found to contribute significantly to the response level, despite their heavy weightings in the resulting prediction equation. When multiple observations were made at each of the data points, however, the subsequent analysis revealed that all of the experimental variables were significant in determining the response level. Of course, when the basic central-composite design is modified in such a manner, methodological questions arise concerning how best to retain the positive attributes of

the basic design, while still making the modifications appropriate to research with human subjects. For example, should repeated observations be made at more than one experimental point; should all data be retained or should they be collapsed; should different subjects be observed at each experimental point or should the same subjects be observed all points; under what conditions are particular design variations especially appropriate?

The following discussion proposes several design variations appropriate to human factors research together with the ensuing methodological considerations. A generalized computer program to analyze data from each of these design variations as well as data from the basic central-composite design has been developed by Clark, Williges, and Carmer (1971).

Collapsed Designs

The simplest modification is achieved merely by replicating the entire central-composite design a given number of times. Consider, for example, the orthogonally blocked, central-composite design depicted in Figure 2. Suppose the investigator elects to replicate that design five times. The data points remain the same as those listed under Figure 2. Now, however, the design involves a total of 100 observations, over a total of 15 treatment conditions. Block 1 now contains 30 observations, Block 2 contains 30 observations, and Block 3 contains 40 observations. Note that, although multiple observations have been made at each of the experimental points, the center point has still been replicated six times more than any other point. While the points on the surface of the sphere have been replicated five times, the center point has been replicated 30 times, 10 times within each of the three blocks.

At this point the investigator must decide whether or not to retain and analyze directly the data corresponding to all 100 observations. He could collapse his data across those subjects within the same block who were observed at the same experimental point and, then analyze the collapsed data. The net effect of collapsing in this manner is a data matrix identical in form and number of observations to one resulting from the original blocked central-composite design shown in

Figure 2. Now, however, the data are combined values obtained from collapsing rather than data representing a simple observation. In addition, estimates of experimental error are obtained from the resulting six center points, each of which is a collapsed score.

This procedure has been used successfully by Mueller and Simon (1969) in a target recognition experiment. It has the advantage of retaining all the features of a RSM central-composite design as well as adding stability to the experimental data points because the collapsed data are not heavily biased by the results of any one extreme subject. This is especially valid if the median is used as the combining statistic. Since it is probably of little value to develop unique prediction equations for each subject, such a collapsing procedure may be appropriate even though degrees of freedom are lost from the design.

Noncollapsed Designs

Suppose, however, that the investigator replicating the blocked central-composite design chooses not to collapse his data across subjects. Rather, he retains each of the subject's data for subsequent analysis. By retaining all this information he gains degrees of freedom for the error term which were previously lost by collapsing the data. Error is now estimated across all points at which replications occur, instead of using only the estimate of the error at the center point. It is quite possible that there may be certain areas of the experimental region in which there is considerable variability in response and other areas in which the variability is negligible. This is particularly true if the range of factor levels under consideration is sizable. Given this variability, it is not reasonable to use the estimate of error at only one such area as an estimate of error throughout the experimental region. The prediction equation which one develops should afford a reasonable description of the entire response surface, not merely a selected area of that response surface.

When noncollapsed designs are used, the investigator must make another major decision with respect to his selected design. If, due to the nature of his research problem, he chooses to observe different subjects at each of the experimental

points, the resulting study constitutes a between-subjects design. If, on the other hand, he elects to observe each of a set of subjects under all experimental conditions, the resulting study constitutes a within-subject design. The choice of a between- versus a within-subject design is dictated by the particular question which the researcher is investigating. In either case, if the necessary restrictions are observed, the design conforms to the basic central-composite design.

Between-Subjects Designs. Given certain research questions, observing the same subjects under more than one experimental condition would lead one to draw invalid conclusions concerning the effect of the various experimental manipulations. Consider, for example, an investigation of the comparative efficacy of selected training methods. Certainly Training Method B cannot be evaluated accurately by observing the performance of subjects who have previously been trained to criterion under Method A, because the observed performance may be a function of not only the condition itself, but also of the preceding condition which he has experienced. In such a case it is imperative that the investigator adopt a between-subjects design, observing each subject under only one experimental condition.

Recall the detection latency study which replicates the orthogonally blocked central-composite design of Figure 2 five times. If 100 different subjects are observed across those 20 data points (six of which are the center point), a between-subjects design is realized. Since the full central-composite design is being replicated in tact, the necessary relationship guaranteeing orthogonal blocking, as given in Equation 1, is still satisfied. As in the original design the center point is being replicated six times more than any other point. Although experimental error is now being estimated across all data points, the results of a subsequent analysis to determine a first order prediction equation are of the same type shown in Table 2. The increased number of observations is reflected in the values for total degrees of freedom, residual degrees of freedom, and error degrees of freedom; the adjusted values are 99, 96, and 83, respectively. Meyer (1963) has used this design procedure successfully in a human learning experiment.

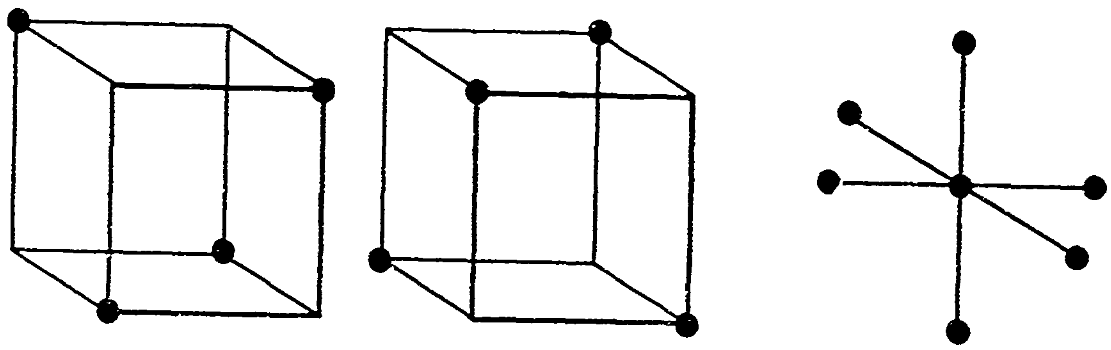
If, indeed, the variability in response at each of a series of data points is used as an estimate of experimental error variance, there is no need to replicate

one point more than any other. In the original central-composite design, in which only the center point is replicated, the additional observations at that point provide the investigator with his only estimate of error. But, with repeated observations occurring at each of the experimental points, there appears no need to make more observations at the center. The investigator could choose instead to replicate each of the experimental points, including the center, an equal number of times, while still maintaining the use of different subjects for each observation.

Eliminating observations at the center point, however, has implications for orthogonal blocking. It is now necessary to adjust the value of α accordingly, because the original blocking has been disturbed due to the elimination of center points from the factorial portion of the design and the reduction in the number of center points in the star portion of the design. With respect to the target detection latency example in which repeated observations are made at each of 15 unique experimental points, making the appropriate adjustment results in an α value of 1.87 rather than 1.633, as defined by Equation 1. This change in the α value is reflected in Figure 3 which designates the orthogonal blocking of the 15 unique experimental points. Note the reduction of data collection points within each of the three blocks, and the complete absence of center points in Blocks 1 and 2.

Changing the coded value of α also necessitates reselecting the real world levels of the various factors under study. Recalculating the levels of display resolution, for example, the investigator learns that the five levels appropriate to the new orthogonally blocked design are 149, 236, 336, 436, and 523. Selecting these five levels retains the center of the experimental region, but increases its range over that indicated in Figure 2.

Replicating this modified central-composite design five times, the investigator makes a total of 75 observations, 20 in Block 1, 20 in Block 2, and 35 in Block 3. Submitting these 75 observations to direct analysis to determine a first-order prediction equation yields results similar to those shown in Table 2. Again, the change in design is reflected in corresponding changes in values of total degrees of freedom, residual degrees of freedom, and error degrees of freedom; the adjusted values are 74, 71, and 60, respectively.



BLOCK 1	BLOCK 2	BLOCK 3
(1, -1, 1)	(-1, 1, 1)	(-1.87, 0, 0)
(1, 1, -1)	(-1, -1, 1)	(0, -1.87, 0)
(-1, 1, 1)	(1, -1, -1)	(0, 0, -1.87)
(-1, -1, -1)	(1, 1, 1)	(1.87, 0, 0)
		(0, 1.87, 0)
		(0, 0, 1.87)
		(0, 0, 0)

Figure 3. Orthogonal blocking of second-order central-composite design in three variables with coded value coordinates of data points employing equal number of replications.

Within-Subject Designs. On occasion the objectives of an experiment make it appropriate and desirable to observe each subject in each treatment condition. In such a case, each individual serves as his own control, and between-subject variability does not affect the experimental conditions. Moreover, observing the same set of subjects under each treatment condition affords another obvious advantage over the between-subjects designs in that fewer subjects are needed to conduct the study, albeit one may encounter the familiar problem of subject attrition. Of course, this design strategy is not appropriate when a subject's performance in one condition is affected by prior experience with any of the other conditions. As previously mentioned, a within-subject design is inappropriate for studying differential training effectiveness. However, it could be used effectively to investigate the differential suitability of various display formats to enhance target detection where there is little or no transfer from display to display. When these within-subject designs are used, caution must be exercised to implement the proper counterbalancing so as to avoid spurious order effects.

The within-subject design combines several features of the central-composite design variations previously discussed. Again, a check should be made to insure that the selected value guarantees orthogonality in the case of blocked designs, or rotatability in the case of unblocked designs. The appropriate real world levels of the experimental factors are then determined accordingly. Data are retained, uncollapsed from repeated observations made at each of the experimental points, thereby affording increased degrees of freedom for the resulting error term. As in the other design variations, the within-subject design permits tests for the significance of blocking and of lack of fit as well as tests of individual partial regression coefficients. In addition, a subject term can be isolated and tested for significance. Because subjects are completely crossed with treatments (every subject receives every treatment once), one can refine his estimate of experimental error variance by accounting for the variability within the individual subjects after assessing the variability within treatment conditions. The error term which results from merely accounting for the variability of response at the experimental points

irrespective of subject is confounded by an interaction between subjects and the treatment conditions. By removing this effect, a better estimate of experimental error is available for subsequent tests for significance of the partial regression weights of the RSM prediction equations.

By way of example, the same four subjects might be observed at each of the 15 experimental points designated in Figure 3, thereby yielding a total of 60 observations. Hypothetical data for such a design are presented in Table 3. Note that the 1.87 value for α is still appropriate because all 15 points, including the center point, are being replicated an equal number of times. A multiple regression analysis of these hypothetical data yields the following first order prediction equation:

$$Y = 16.44 - 1.16751591\ X_1 - 0.39631381\ X_2 + 0.82118942\ X_3$$

Substituting given levels of display resolution, visual angle, and random noise for X_1, X_2, and X_3, respectively, into this equation provides a corresponding predicted level of detection latency.

The results of a subsequent ANOVA performed on the hypothetical data of the regression analysis appear in Table 4. Note the additional "subjects" component into which residual variance has been subdivided. The corresponding degrees of freedom reflect the use of four subjects throughout the experiment. Notice also that the error degrees of freedom are reduced by 3, the degrees of freedom attributed to the subject factor. Had this experiment utilized different subjects throughout, the value of error degrees of freedom would have been 45 rather than 42. But, in the case of within-subject designs, the error term is purified by removing the subject effect from it.

CONCLUSIONS

The techniques of RSM, and the central-composite design in particular, can be effectively used in human factors research, where the goal is frequently the development of an equation to describe the relationship between human performance and a host of equipment parameters. Certain modifications in the basic

TABLE 3

Hypothetical Data in Coded Form For A Three-Factor, Central-Composite, RSM Design Using Repeated Measures on Four Subjects

			Detection Latency (Seconds) For Four Subjects			
Resolution	Visual Angle	Random Noise	S_1	S_2	S_3	S_4
1.0	-1.0	1.0	15.8	15.9	16.1	16.4
1.0	1.0	-1.0	14.3	14.5	14.0	14.8
-1.0	1.0	1.0	17.0	17.3	17.1	16.9
-1.0	-1.0	-1.0	17.4	17.5	17.0	17.3
-1.0	1.0	-1.0	16.8	16.7	17.0	17.0
-1.0	-1.0	1.0	18.1	18.3	18.6	18.1
1.0	-1.0	-1.0	14.9	15.2	14.5	15.0
1.0	1.0	1.0	16.2	16.7	16.4	15.9
-1.87	0.0	0.0	19.0	19.1	18.9	19.5
0.0	-1.87	0.0	17.3	16.9	17.4	16.8
0.0	0.0	-1.87	15.1	15.3	14.4	15.0
1.87	0.0	0.0	13.9	14.2	13.7	14.1
0.0	1.87	0.0	14.9	15.0	14.8	15.0
0.0	0.0	1.87	19.2	19.0	20.0	18.9
0.0	0.0	0.0	15.8	16.1	16.4	16.0

TABLE 4

First-Order Regression Analysis of Variance Summary Table
for Hypothetical Detection Latency Data of Four Subjects

Source	Percent	df	Sum of Squares	Mean Square	F-Value
Regression	(84.98)	(3)	131.62	43.87	548.37 ***
b_1	52.78	1	81.75	81.75	1021.87 ***
b_2	6.08	1	9.42	9.42	117.75 ***
b_3	26.11	1	40.44	40.44	505.50 ***
Residual	(15.02)	(56)	23.27	0.42	
Blocks	0.83	2	1.29	0.65	8.13 **
Subjects	0.09	3	0.14	0.05	0.63
Lack of Fit	11.84	9	18.34	2.04	25.50 ***
Error	2.26	42	3.50	0.08	
Total	100.00	59	154.88		

Multiple Regression Coefficient = 0.92

Coefficient of Determination = 0.85

** $p < .01$

*** $p < .001$

central-composite design, however, appear to make the method more appropriate to research involving human subjects. In making the appropriate design modifications, the investigator must make several major decisions. He must decide whether or not to make repeated observations over a series of experimental points rather than at a single point. If his goal is to develop a global prediction equation to approximate the entire response surface, replications at each of the experimental data-collection points appears to be a wise strategy. The basic central-composite design, calling for replications at only the center point, is perhaps better reserved for preliminary research where the primary aim is to ascertain quickly what major factors appear worthy of more thorough study.

The investigator must also select either a between-subjects or a within-subject design. This choice is dictated by the objectives of his particular experiment. Of the design variants discussed above, those advocating multiple and equal replications at all experimental points, followed by analysis of uncollapsed data, appear the most advantageous, whether they are conceived as between- or within-subject designs. The particular modifications which the investigator elects to implement have ramifications for other aspects of the design such as orthogonal blocking and rotatability. Appropriate adjustments must be made in factor level selection in order to retain such attributes in view of the overall design modification.

REFERENCES

Baron, M. L. and Williges, R. C. Transfer of training assessment by means of response surface methodology. Paper presented at the Human Factors Society annual meeting, October 1971.

Box, G. E. P. and Hunter, J. S. Multifactor experimental designs for exploring response surfaces. Annals of Mathematical Statistics, 1957, 28, 195-241.

Box, G. E. P. and Wilson, K. B. On the experimental attainment of optimum conditions. Journal of the Royal Statistical Society, Series B Methodological, 1951, 13, 1-45.

Clark, Christine, Williges, R. C., and Carmer, S. G. General computer program for response surface methodology analyses. Savoy, Illinois: University of Illinois, Institute of Aviation, Aviation Research Laboratory, Technical Report ARL-71-8/AFOSR-71-1, May 1971.

Cochran, W. G. and Cox, Gertrude M. Experimental designs. Chapter 8A. Some methods for the study of response surfaces. (2nd ed.) New York: Wiley, 1957, 335-375.

Meyer, D. L. Response surface methodology in education and psychology. The Journal of Experimental Education, 1963, 31, 329-336.

Mueller, B. H. and Simon, C. W. Evaluation of infrared video high speed computation. Wright-Patterson Air Force Base, Ohio: Air Force Avionics Laboratory, Technical Report AFAL-TR-69-48, March 1969.

Simon, C. W. The use of central-composite designs in human factors engineering experiments. Culver City, California: Hughes Aircraft Co., Display Systems and Human Factors Department, Technical Report AFOSR-70-6, December 1970.

Williges, R. C. and Simon, C. W. Applying response surface methodology to problems of target acquisition. Human Factors, 1971, in press.